BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 8: e133089
doi: 10.3897/biss.8.133089

OPEN ACCESS

Conference Abstract

Bridging Data Standards and FAIR Principles in Biodiversity Digital Twinning: Prototyping, Challenges, Lessons Learned, and Future Plans

Sharif Islam[‡,§], Julian Lopez Gordillo[‡], Dag Endresen[|], Carrie Andrew[|]

‡ Naturalis Biodiversity Center, Leiden, Netherlands
§ DiSSCo, Leiden, Netherlands
| Natural History Museum, University of Oslo, Oslo, Norway

Corresponding author: Sharif Islam (sharif.islam@naturalis.nl)

Received: 25 Jul 2024 | Published: 25 Jul 2024

Citation: Islam S, Lopez Gordillo J, Endresen D, Andrew C (2024) Bridging Data Standards and FAIR Principles in Biodiversity Digital Twinning: Prototyping, Challenges, Lessons Learned, and Future Plans. Biodiversity Information Science and Standards 8: e133089. https://doi.org/10.3897/biss.8.133089

Abstract

Digital twins combine modelling, domain knowledge, computing power, and multiple datasets to offer the potential to unlock new insights into biodiversity (de Koning et al. 2023). The Biodiversity Digital Twin (BioDT) project pioneers this approach to aid in understanding biodiversity through prototyping digital twins (Golivets et al. 2024). However, working with biodiversity data presents challenges due to their dynamic and diverse nature as well as the need for having to deal with incompleteness, uncertainties (Rocchini et al. 2011), disproportionate representation patterns in global studies from wealthier economies (Hughes et al. 2024), and issues with data aggregation and integration (Wüest et al. 2020). Similar to BioDT, there are also plans for creating a Digital Twin of the Ocean (DTO). DTO-Bioflow project addresses these data challenges in the marine domain, where studies show that although European seas host 48,000 marine species (75% described), the data are not yet FAIR (Findable, Accessible, Interoperable, and Reusable) (Ramírez et al. 2022). These challenges hamper the modelling capabilities needed for effective predictions and conservation prioritisation.

The adaptability of digital twins across temporal and spatial scales and their ability to model dynamic ecosystems make them ideal for biodiversity research and real-time

© Islam S et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

conservation efforts. However, their success hinges on the consistent integration and alignment of data from disparate sources (Trantas et al. 2023). This integration involves standardising terms used to describe datasets, such as temporal coverage or controlled vocabularies like "Forest" for targetHabitatScope. Thus, adopting data standards is essential. Additionally, challenges such as model bias (Lewers 2023), research context, and data provenance must be considered, adding complexity to metadata capture and alignment.

BioDT addresses these challenges with modular building blocks for data integration, model deployment, and workflow management. This approach facilitates the gradual adoption of data standards and FAIR principles, which need to encompass not just data, but also models and software. As automation, ease of reproducibility, and deployability are critical for digital twinning success, data integration and interoperability issues may arise due to missing or insufficient parameter descriptions in the model, incomplete information on data selection, and the unavailability of required software package details. Thus, data standardisation provides a pathway for a consistent approach that can be adopted for different use cases.

Common data sources in BioDT, like species occurrences and environmental variables, benefit from standards such as Darwin Core (Wieczorek et al. 2012) and the Ecological Metadata Language (Jones et al. 2019). While valuable, these standards may not fully encompass the complexity needed for comprehensive biodiversity digital twins. Additionally, differing familiarity with these standards and FAIR principles among communities pose challenges. Continuous adoption of data standards, alongside exploring complementary approaches like schema.org or bioschemas.org for capturing diverse (meta)data, is essential. Collaboration with data providers, modellers, and various research infrastructures is also crucial (Andrew et al. 2024).

We share our experience using Research Object Crate (RO-Crate), leveraging common JavaScript Object Notation for Linked Data (JSON-LD) representation for metadata profiles and workflow representation, to connect with different infrastructures. In the BioDT project, we are working with various use cases to create prototype digital twins that can serve as valuable resources for other projects. The evolving landscape of digital twin concepts, along with other European Union-funded initiatives like DTO-Bioflow and Destination Earth (DestinE), emphasises the importance of alignment within the digital twin ecosystem. BioDT is committed to aligning with and contributing to this broader context, highlighting the critical role of data standardisation and FAIR implementation.

Keywords

standards mapping, metadata alignment, data integration, standards adoption

Presenting author

Sharif Islam

Presented at

SPNHC-TDWG 2024

Funding program

HORIZON.1.3 - Research infrastructures

HORIZON-MISS-2022-OCEAN-01-07 - Integration of biodiversity monitoring data into the Digital Twin Ocean

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Andrew C, Islam S, Weiland C, Endresen D (2024) Biodiversity data standards for the organization and dissemination of complex research projects and digital twins: a guide. arXiv 2405.19857 https://doi.org/10.48550/arXiv.2405.19857
- de Koning K, Broekhuijsen J, Kühn I, Ovaskainen O, Taubert F, Endresen D, Schigel D, Grimm V (2023) Digital twins: dynamic model-data fusion for ecology. Trends in Ecology & Evolution 38 (10): 916-926. https://doi.org/10.1016/j.tree.2023.04.010
- Golivets M, Islam S, Wohner C, Grimm V, Schigel D (2024) Building Biodiversity Digital Twins: Special Collections. Research Ideas and Outcomes https://doi.org/10.3897/rio.coll.240
- Hughes A, Dorey J, Bossert S, Qiao H, Orr M (2024) Big data, big problems? How to circumvent problems in biodiversity mapping and ensure meaningful results. Ecography https://doi.org/10.1111/ecog.07115
- Jones M, O&rsquo M, Brien BM, Boettiger C, Schildhauer M, Maier M, Whiteaker T, Earl S, Chong S (2019) Ecological Metadata Language version 2.2.0. KNB Data Repository https://doi.org/10.5063/F11834T2
- Lewers K (2023) Leveraging AI in Biodiversity Informatics: Ethics, privacy, and broader impacts. Biodiversity Information Science and Standards 7 https://doi.org/10.3897/biss.7.112701
- Ramírez F, Sbragaglia V, Soacha K, Coll M, Piera J (2022) Challenges for Marine Ecological Assessments: Completeness of Findable, Accessible, Interoperable, and Reusable Biodiversity Data in European Seas. Frontiers in Marine Science 8 https://doi.org/10.3389/fmars.2021.802235

- Rocchini D, Hortal J, Lengyel S, Lobo JM, Jimenez-Valverde A, Ricotta C, Chiarucci A (2011) Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. Progress in Physical Geography 35 (2): 211-226. https://doi.org/10.1177/0309133311399491
- Trantas A, Plug R, Pileggi P, Lazovik E (2023) Digital twin challenges in biodiversity modelling. Ecological Informatics 78 https://doi.org/10.1016/j.ecoinf.2023.102357
- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7 (1). https://doi.org/10.1371/journal.pone.0029715
- Wüest RO, Zimmermann NE, Zurell D, Alexander JM, Fritz SA, Hof C, Karger DN (2020) Macroecology in the age of Big Data-Where to go from here? Journal of Biogeography 47 (1): 1-12. https://doi.org/10.1111/jbi.13633